

徐坤财, 张宁, 廖益龙. 基于 LASSO 回归的脑部胶质肿瘤 MRI 预测方法[J]. 智能计算机与应用, 2024, 14(11): 113–116.  
DOI: 10.20169/j.issn.2095-2163.24051306

# 基于 LASSO 回归的脑部胶质肿瘤 MRI 预测方法

徐坤财, 张宁, 廖益龙

(贵阳信息科技学院 智能工程学院, 贵阳 550025)

**摘要:** 针对临床中不同级别的脑部胶质肿瘤预测准确性低的问题, 本文提出基于 LASSO 回归的脑部胶质肿瘤 MRI 预测方法。首先, 基于 Pyradiomics 提取 BraTS2019 数据集中的 T1ce 增强图像脑部胶质肿瘤的一阶统计特征、形态特征和纹理特征。然后, 采用 Mann-Whitney U 检验对不同级别的脑部胶质肿瘤进行特征选择, 并基于 LASSO 采用 10 折交叉验证构建临床预测模型。最后, 通过对比 SVM 模型和 ELM 模型的诊断效能, 结果显示基于 LASSO 回归的预测模型具有更优的诊断性能。

**关键词:** 脑部胶质肿瘤; 影像特征; 统计特征; 交叉验证; LASSO

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)11-0113-04

## MRI prediction method for brain glioma based on LASSO regression

XU Kuncai, ZHANG Ning, LIAO Yilong

(School of Intelligent Engineering, Guiyang Institute of Information Science and Technology, Guiyang 550025, China)

**Abstract:** Addressing the issue of low accuracy in predicting brain gliomas of different levels in clinical practice, this paper proposes a brain glioma MRI prediction method based on LASSO regression. Firstly, based on Pyradiomics, the first-order statistical features, morphological features, and texture features of brain gliomas in T1ce enhanced images from the BraTS2019 dataset are extracted. Then, Mann Whitney U test is used to select features for different levels of brain gliomas, and a clinical prediction model is constructed using 10 fold cross validation based on LASSO. Finally, by comparing the diagnostic performance of SVM model and ELM model, the results show that the prediction model based on LASSO regression has better diagnostic performance.

**Key words:** brain glioma; imaging features; statistical features; cross validation; LASSO

## 0 引言

胶质瘤是由胶质细胞恶性增生而成的肿瘤。通常发生在大脑、小脑、脑干和脊髓等部位。胶质瘤位于脑组织内、起源于胶质细胞。其发病率约占颅内肿瘤 35.2%~61.0%, 由成胶质细胞衍化而来, 具有发病率高、复发率高、死亡率高以及治愈率低的特点<sup>[1]</sup>。胶质瘤常见症状包括头痛、恶心、呕吐、神经功能障碍、认知障碍、肌肉无力和癫痫发作<sup>[2]</sup>。因此对脑部胶质肿瘤进行诊断具有重要的临床价值。

WHO 根据组织学表现对胶质瘤进行分级<sup>[3]</sup>。具体如下。

(1) Grade I (一级): 是最低级别的胶质瘤, 通常被认为是良性的, 在组织学上, 该肿瘤细胞的外观与正常细胞相似, 通常生长缓慢, 不具有侵袭性。

(2) Grade II (二级): 是低级别的胶质瘤, 有时被称为低级星形细胞瘤或少突细胞瘤。组织学上, 显示出一些异常细胞变化, 但仍与正常细胞相对接近, 生长速度较慢, 但有时可以逐渐恶化。

(3) Grade III (三级): 是高级别的胶质瘤, 如高级星形细胞瘤, 组织学上, 肿瘤细胞显示出更多的异常特征和增殖活性, 具有更高的侵袭性和恶性倾向。

(4) Grade IV (四级): 是最高级别的胶质瘤, 最常见的是胶质母细胞瘤 (Glioblastoma), 在组织学上, 肿瘤细胞异常严重, 具有非常高的增殖活性, 具有强烈

**基金项目:** 贵州省青年科技人才成长项目(黔教技[2024]279号); 贵州省科技计划项目(黔科合支撑[2024]一般178号); 贵州省教育科学规划课题(2024C018)。

**作者简介:** 张宁(1993—), 男, 助教, 主要研究方向: 机器视觉; 廖益龙(1986—), 男, 讲师, 主要研究方向: 机械设计控制。

**通信作者:** 徐坤财(1997—), 男, 助教, 主要研究方向: 类脑计算, 医学图像处理。Email: xkc1009@163.com。

收稿日期: 2024-05-13

的侵袭性,生长速度非常快,且恶性程度很高。

脑胶质瘤的早筛查、早发现、早治疗是与患者提高患者身心健康密切相关的有效途径和方法。临床诊断中,放射科医生通常通过电子计算机断层扫描(Computed Tomography, CT)和MRI影像采用经验性和观察性等具有主观因素的指标进行评估。临床中不同级别的胶质瘤具有不同的治疗方式,对于良性脑胶质瘤,通常采用手术切除的方式进行治疗。恶性脑胶质瘤通常采用放射治疗、化疗和靶向治疗的方式来治疗<sup>[4]</sup>。

最近几年,影像组学在医学图像分析方面得到广泛应用<sup>[5-10]</sup>。影像组学能提取医学图像中肿瘤的高通量特征,从而利用该特征构建辅助诊断模型,为临床诊断提供无创、量化的诊断方案<sup>[11-17]</sup>。其主要原因是提取病灶肿瘤的特征具有解释性强、数据需求较小和计算效率高的优点。因此本文提取了脑部胶质瘤 T1ce 的影像组学特征,并进行特征选择,构建了临床辅助诊断模型,以提高对不同级别的脑部胶质瘤分类准确率。

## 1 材料与方法

### 1.1 数据集与预处理

本文利用脑部肿瘤数据集 BraTS2019 进行实验分析,该数据集包含了 259 例高级别胶质肿瘤(High Grade Gliomas, HGG)和 76 例低级别胶质肿瘤(Low Grade Gliomas, LGG),每例患者包含 4 种模态的 MRI 数据,分别为 T1、T2、T1ce 和 Flair。每张 MRI 图像的大小为 240 像素×240 像素×155 像素,每种模态的图像数据均保存为 nifti 文件(.nii.gz),其中手动标注结果包含 4 个标签:标签 0 为背景,标签 1 为坏死核(Necrotic Core),标签 2 为水肿(Edema),标签 4 为增强核(Enhancing Core)。数据集地址为 <http://braintumorsegmentation.org/>。

本文使用 T1ce 数据进行实验,并且将标注结果为 2 和 4 的重新创建为标签 1,标签 0 保持不变。数据按照 7 : 3 随机划分为训练集和测试集,则训练集 236 例,包含 182 例 HGG、54 例 LGG;测试集 99 例,包含 77 例 HGG、22 例 LGG。数据分布情况见表 1。

表 1 脑部肿瘤数据分布情况

Table 1 Distribution of brain tumor data

数据集	HGG	LGG	例
训练集	182	54	
测试集	77	22	

### 1.2 特征提取

本文使用 Pyradiomics 库(版本:3.0)提取肿瘤感兴趣区域内的影像特征,而在本次特征提取过程中 *normalize* 为 True, *normalizeScale* 为 500, *resampledPixelSpacing* 为 [1,1,1], *padDistance* 为 10。分别提取了一阶统计特征、形态特征、二阶及高阶纹理特征。其中,一阶统计特征反映所测体素的对称性、均匀性以及局部强度分布变化;形态特征是描述肿瘤三维信息的大小和形态;二阶及高阶纹理特征反映了图像体素灰度之间的空间排列关系,主要分为灰度共生矩阵(GrayLevelCooccurrence Matrix, GLCM)、灰度游程长度矩阵(Gray LevelRun Length Matrix, GLRLM)、灰度尺寸区域矩阵(Gray LevelSize Zone Matrix, GLSZM)、邻域灰度差异矩阵(Neighbouring Gray Tone Difference Matrix, NGTDM)和灰度相关矩阵(Gray Level Dependence Matrix, GLDM)。共提取了 107 个影像特征。

### 1.3 特征选择

对于提取的 107 个影像特征,采用 Mann-Whitney U Test 对训练集中 HGG 和 LGG 进行统计分析,选择 HGG 与 LGG 之间具有统计差异( $P < 0.05$ )的影像特征。通过 Mann-Whitney U Test 选择之后存在 HGG 与 LGG 之间有 92 个影像特征具有统计意义。

### 1.4 模型构建

对于特征选择之后的 92 个影像特征,采用最小绝对收缩和选择算子(Least Absolute Shrinkage and Selection Operator, LASSO)<sup>[18]</sup>进一步选择特征并构建模型,最终选择了 6 个非零系数的影像特征参与构建了预测模型,模型构建过程中采用了 10 折交叉验证,选择的 6 个特征中包含了 2 个一阶统计特征(RootMeanSquared、Skewness)和 4 个二阶及高阶纹理特征(GLDM\_LargeDependenceLowGrayLevelEmphasis(GLDM\_LDLGLE), GLSZM\_LargeAreaEmphasis(GLSZM\_LAE), GLSZM\_SmallAreaEmphasis(GLSZM\_SAE), GLSZM\_ZoneVariance(GLSZM\_ZV))。并基于 6 个特征计算了每例患者的预测得分。模型构建过程可视化如图 1 所示。预测得分,即模型构建的计算公式为:

$$\begin{aligned}
 Score = & 1.608 + 0.573 \times RootMeanSquared + \\
 & 0.385 \times Skewness - 0.012 \times \\
 & GLDM\_LDLGLE - 0.004 \times GLSZM\_LAE + \\
 & 0.808 \times GLSZM\_SAE - 0.001 \times \\
 & GLSZM\_ZV
 \end{aligned} \quad (1)$$

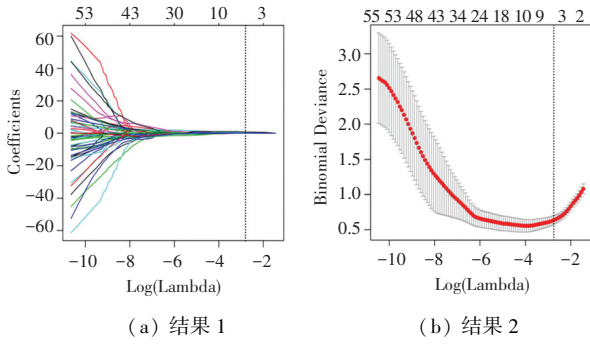


图 1 LASSO 模型构建可视化

Fig. 1 Visualization of LASSO model construction

1.5 统计分析和评价指标

采用 Python (版本:3.6) 提取影像组学特征、使用 R 语言 (版本:3.4.4, <https://www.r-project.org/>) 的 “pROC” 绘制受试者工作特征曲线 (receiver operating characteristic curve, ROC), 使用 Matlab2016b 版本中进行 Mann-Whitney U Test 选择特征。采用曲线下面积 (Area Under Curve, AUC)、Accuracy、灵敏度、特异度、PPV 和 NPV 等指标评价模型的预测性能。采用

综合判别改善指数 (Integrated Discrimination Improvement, IDI) 来评价模型之间的改善能力。IDI 大于 0, 表示新模型相对于旧模型具有正改善能力; IDI 等于 0, 表示新模型没有改善能力; IDI 小于 0, 表示新模型的诊断不如旧模型。

2 结果与分析

为了对比基于 LASSO 模型的预测性能, 本文基于 Mann-Whitney U Test 之后的 92 个影像特征构建了 ELM 预测模型和 SVM 预测模型<sup>[19-20]</sup>。并且在测试集上进行测试性能, 实验结果见表 2。LASSO 模型、SVM 模型和 ELM 模型在训练集上的曲线下面积 AUC 分别为 0.934 (95% CI: 0.894~0.976)、0.956 (95% CI: 0.926~0.987) 和 0.940 (95% CI: 0.906~0.973); 测试集上的曲线下面积 AUC 分别为 0.956 (95% CI: 0.926~0.997)、0.927 (95% CI: 0.871~0.983) 和 0.924 (95% CI: 0.870~0.978)。对应的 ROC 曲线如图 2 所示。

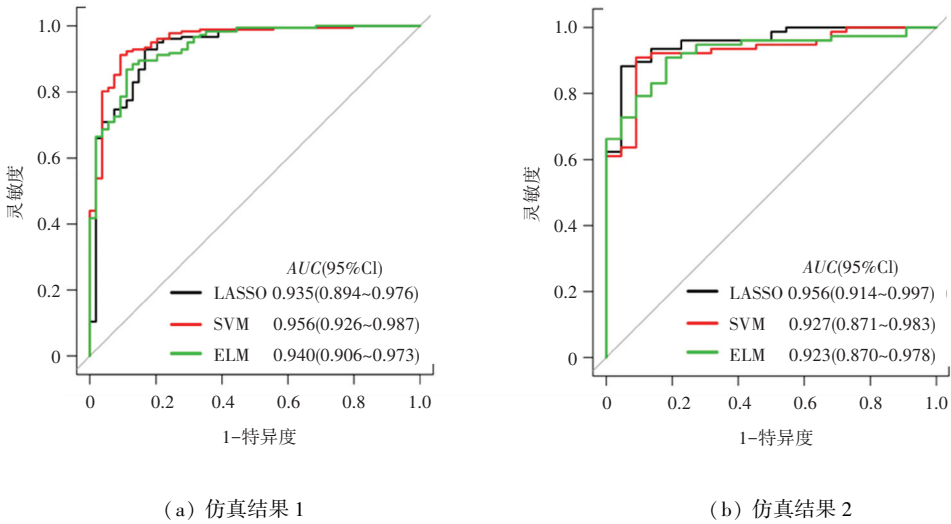


图 2 不同分类模型在训练集和测试集上的 ROC 曲线

Fig. 2 ROC curves of different classification models on training sets and test sets

表 2 不同分类模型的对脑部肿瘤 HGG 和 LGG 的诊断性能

Table 2 Diagnostic performance of different classification models for High-Grade Glioma (HGG) and Low-Grade Glioma (LGG) in brain tumors

Model	Data	AUC (95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
LASSO	Train	0.935 (0.894~0.976)	0.907	0.929	0.833	0.949	0.776
	Test	0.956 (0.914~0.997)	0.899	0.883	0.955	0.986	0.700
SVM	Train	0.956 (0.926~0.987)	0.911	0.912	0.907	0.971	0.754
	Test	0.927 (0.871~0.983)	0.909	0.909	0.909	0.972	0.741
ELM	Train	0.940 (0.906~0.973)	0.873	0.868	0.889	0.963	0.667
	Test	0.924 (0.870~0.978)	0.889	0.909	0.818	0.946	0.720

使用 IDI 做进一步对比, LASSO 模型均优于 SVM 模型和 ELM 模型, 在测试集上 LASSO 模型相比于 SVM 模型和 ELM 模型的 IDI 值分别为  $IDI = 0.1136$  ( $P = 0.0033$ ) 和  $IDI = 0.1558$  ( $P = 0.0004$ )。IDI 结果显示, 组合模型具有更好的诊断性能。

### 3 讨论

基于 T1ce 增强图像的影像特征构建的诊断模型能较好地预测不同级别的脑部胶质肿瘤。本文的研究表明采用基于 LASSO 进行二次特征选择并构建的诊断预测模型的诊断性能最高, 在测试集上的 AUC 达到 0.956。能为临床诊断提供定量化的参考。

在提取的 107 个影像特征中, 最后选择了 6 个特征构建预测模型, 该特征能反映肿瘤体素之间的变化。其中, *RootMeanSquared* 表示肿瘤体素的均方根误差; *Skewness* 表示肿瘤体素测量值的分布关于平均值的不对称性; *GLDM\_LDLGLE* 反映肿瘤体素测量具有较低灰度值的大相关性的联合分布、*GLSZM\_LAE* 反映肿瘤大面积区域分布的度量, 值越大、区域越大, 纹理表现越粗糙; *GLSZM\_SAE* 反映肿瘤小尺寸区域分布的度量, 数值越大、区域越小, 纹理表现越精细; *GLSZM\_ZV* 反映肿瘤像素区域方差。上述 6 个特征反映了肿瘤不同维度的表现, 并且基于该特征能够有效区分脑部不同级别的胶质肿瘤。

综上所述, 本研究提取 T1ce 增强图像的影像特征, 并进行特征选择, 采用 LASSO 进行 10 折交叉验证构建预测模型。与 SVM 模型和 ELM 模型相比, 该模型具有简单稳定、计算复杂度低、泛化性能高等优点。

本研究仍有一定的局限性:

(1) 本研究中病灶分割采用手动勾画完成, 并且存在一定的主观性, 该过程中费时费力。

(2) 本研究属于回顾性研究, 在样本选择中难免存在一定的偏倚。

在今后的研究中, 将增加病例样本数, 并采用先进的图像分割技术来对脑部胶质瘤的图像展开研究和分析。

### 4 结束语

本文提取 T1ce 增强图像影像特征构建基于

LASSO 的脑部胶质瘤预测模型, 该模型与其他预测模式相比具有较高的诊断性能, 可为临床提供一定的参考价值。

### 参考文献

- [1] 宋静, 宗会迁, 张娅, 等. 磁共振波谱联合减影技术在高级别胶质瘤影像组学分级预测的研究[J]. 磁共振成像, 2023, 14(6): 59-65.
- [2] 葛鑫, 刘光耀, 甘铁军, 等. 合成 MRI 联合弥散加权成像评估胶质瘤级别及肿瘤细胞增殖活性[J]. 中国医学影像技术, 2023, 39(2): 171-175.
- [3] 孙崇然, 许晶虹, 张布衣, 等. 2021 年世界卫生组织中枢神经系统肿瘤分类(第五版)儿童型弥漫性胶质瘤分类解读[J]. 中国现代神经疾病杂志, 2021, 21(9): 791-803.
- [4] 张为. 脑胶质瘤化疗耐药机制的研究进展[J]. 中国临床神经外科杂志, 2022, 27(2): 132-134.
- [5] 谢小雪, 金和坤, 肖翔, 等. 影像组学在肿瘤放射治疗学中的研究进展[J]. 中国肿瘤外科杂志, 2022, 14(6): 527-532.
- [6] 周晓媛, 尚圣捷, 王颖妮, 等. 基于瘤内异质性分割的软组织肿瘤 MRI 影像组学辅助诊断研究[J]. 中国医疗设备, 2021, 36(9): 86-90.
- [7] 许珂, 张勇, 程敬亮, 等. 影像组学在中枢神经系统的应用进展[J]. 中华放射学杂志, 2018, 52(10): 805-808.
- [8] 方梓涛, 刘丹, 吴扬东, 等. 融合残差结构和注意力机制的下颌骨 CT 图像分割方法[J]. 智能计算机与应用, 2024, 14(2): 83-89.
- [9] 刘张, 郭旭东, 张璐璐. 基于 EfficientNet 的胶囊内镜图像多病灶协同检测[J]. 智能计算机与应用, 2021, 11(11): 157-161.
- [10] 杨春静. 基于颜色纹理特征的医学图像智能分割算法[J]. 智能计算机与应用, 2023, 13(8): 201-204.
- [11] 陈子昂, 刘娜, 袁野, 等. 基于多重注意力机制的多模态脑肿瘤图像分割[J]. 智能计算机与应用, 2023, 13(12): 80-86.
- [12] 毋小省, 杨奇鸿, 唐朝生, 等. 融合注意力机制的多模态脑肿瘤 MR 图像分割[J]. 计算机辅助设计与图形学学报, 2023, 35(9): 1429-1438.
- [13] 陈柏年, 韩雨童, 何涛, 等. 基于级联动态注意力 U-Net 的脑肿瘤分割方法[J]. 计算机科学, 2023, 50(S2): 1031-1037.
- [14] 宋建丽, 吕晓琪, 谷宇. 语义流引导采样结合注意力机制的脑肿瘤图像分割[J]. 光学精密工程, 2024, 32(4): 565-577.
- [15] 孙康康, 陈伟, 李奇轩, 等. 多模态 MRI 脑肿瘤图像分割方法研究进展[J]. 磁共振成像, 2023, 14(11): 164-169.
- [16] 许学添, 李伶俐, 蔡跃新. 基于多尺度注意力的特征自适应聚合脑肿瘤图像分割[J]. 计算机测量与控制, 2023, 31(12): 224-230.
- [17] 帖军, 朱祖桐, 郑禄, 等. 基于混合空洞卷积与特征融合的肝脏肿瘤图像分割[J]. 电子测量技术, 2023, 46(22): 122-130.
- [18] 倪新洁, 梁彪, 倪佩可. 结合 LASSO 算法与 logistic 回归模型的 P2P 信贷审批结果研究[J]. 统计与管理, 2015(8): 44-47.
- [19] 方欲晓, 何可人. 基于节点路径重构和 ELM 的无线网络 DDoS 攻击源追踪[J]. 现代电子技术, 2024, 47(13): 93-96.
- [20] 赵冬梅, 孙明伟, 宿梦月, 等. 基于改进 SKNet-SVM 的网络安全态势评估[J]. 应用科学学报, 2024, 42(2): 334-349.